Ling 566
Feb 11, 2019

Grammar and Processing

© 2003 CSLI Publications

Overview

- Psycholinguistics and grammar design
  - What grammar has to say
  - What psychological evidence has to say
    - Acquisition
    - Production
    - Comprehension
- Universals

© 2003 CSLI Publications

What does grammar have to do with psychology?

Three ways it could be relevant:

- It provides insight into how children acquire language.
- It provides insight into how speakers produce utterances.
- It provides insight into how listeners understand utterances.

© 2003 CSLI Publications

Our model: Key characteristics

- Surface-oriented
- Constraint-based
- Lexicalist

© 2003 CSLI Publications

Chomsky's position:

- Grammar represents knowledge of language ("competence").
- This is distinct from use of language ("performance").
- We can draw a strong conclusion about language acquisition, namely, most grammatical knowledge is innate and task-specific.
- Serious study of language use (production and comprehension) depends on having a well-developed theory of competence.

© 2003 CSLI Publications

Brief remarks on language acquisition

- Chomsky's nativism is very controversial
  - It is based on the "poverty of the stimulus" argument, and a model of learning as hypothesis testing.
  - The environment may be more informative than he assumes.
  - There may be more powerful learning methods than he assumes.
- There has not been much work on language acquisition using constraint-based lexicalist theories like ours; **but**
  - Explicit formulation is a prerequisite for testing learning models
  - Our feature structures could model richer context information.
- We're neutral with respect to this controversy.

© 2003 CSLI Publications

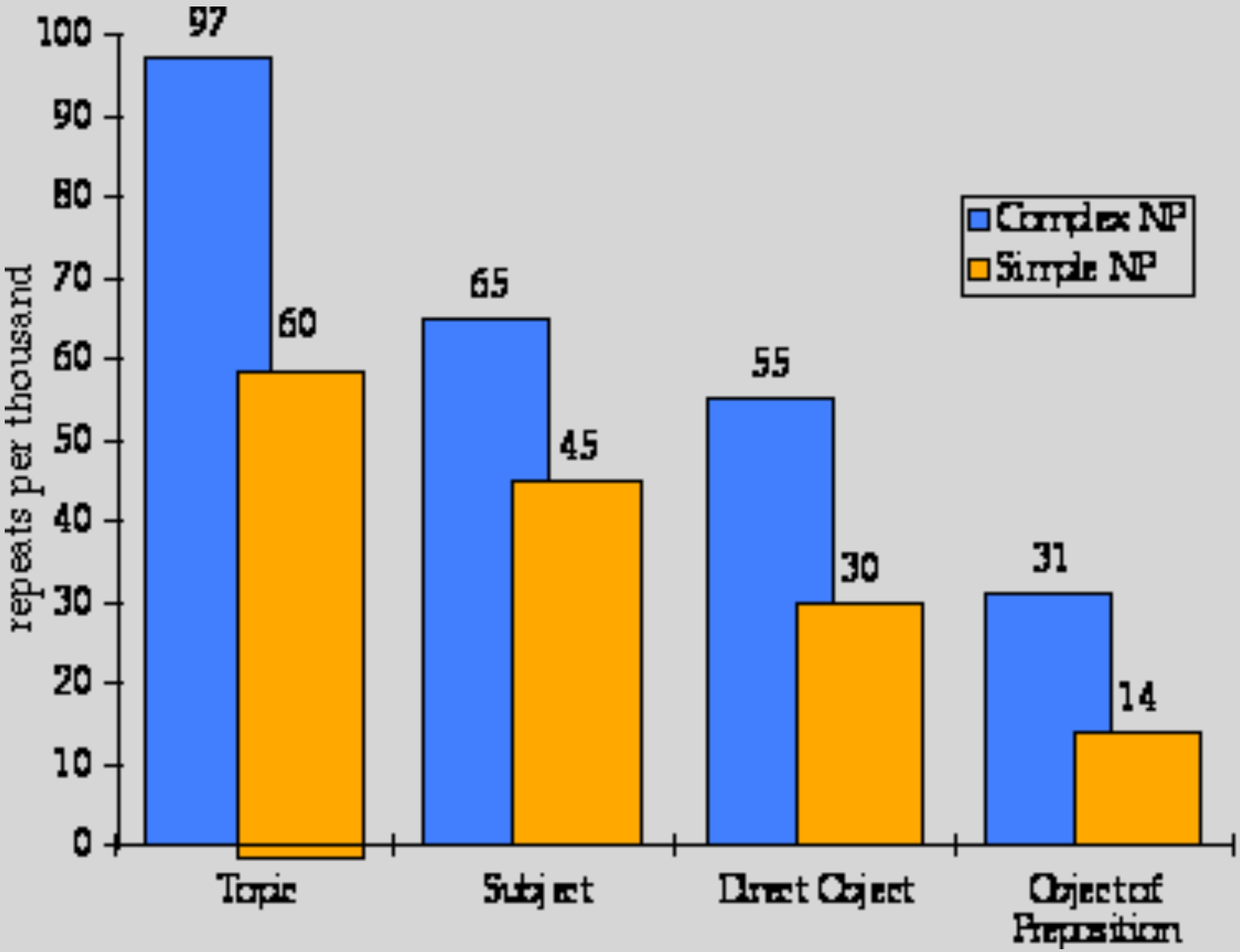
Production and Grammar

- Evidence for left-to-right effects
- Evidence for grammar in processing
- Evidence for top-down planning

© 2003 CSLI Publications

Disfluencies are sensitive to structure:

Repeat rate of *the* varies with position and complexity of the NP it introduces:



© 2003 CSLI Publications

Production errors are sensitive to syntactic structure

Agreement errors are more common with PP complements than sentential complements: errors like (2) are significantly more common than errors like (1).

(1) **The claim that the wolves had raised the babies were rejected*.

vs.

(2) **The claim about the newborn babies were rejected*.

© 2003 CSLI Publications

So why?

- Speculation: Clauses are their own agreement domains, so people don't mistake an NP in a lower clause as a trigger for agreement
- Original work: Kay Bock (1980s).

© 2003 CSLI Publications

Some high-level sentence planning is necessary, too

- *Ich habe* ***dem*** *Mann,* ***den*** *ich gesehen habe geholfen*.

  I have the-dat man who-acc I seen have helped

  "I helped the man I saw"

- *Ich habe* ***den*** *Mann,* ***dem*** *ich geholfen habe gesehen*.

  I have the-acc man who-dat I helped have seen.

  "I saw the man I helped "

- The choice between *dem* and *den* depends on the choice of verbs several words later.

© 2003 CSLI Publications

A production model should allow interaction of top-down and left-to-right information

- Grammar plays a role in production.
- Partial grammatical information should be accessible by the production mechanism as needed.
- This argues against grammatical theories that involve sequential derivations with fixed ordering.
- Our theory of grammar has the requisite flexibility.

© 2003 CSLI Publications

Comprehension

- Early work tried to use transformational grammar in modeling comprehension
- The Derivational Theory of Complexity: The psychological complexity of a sentence increases with the number of transformations involved in its derivation.
- Initial results seemed promising, but later work falsified the DTC.

© 2003 CSLI Publications

Some relevant quotes

- "The results show a remarkable correlation of amount of memory and number of transformations"
  – Chomsky, 1968

- "[I]nvestigations of DTC…have generally proved equivocal. This argues against the occurrence of grammatical derivations in the computations involved in sentence recognition"
  – Fodor, Bever, & Garrett, 1974

© 2003 CSLI Publications

Another quote

- "Experimental investigations of the psychological reality of linguistic structural descriptions have…proved quite successful."
  – Fodor, Bever, & Garrett, 1974

- In particular, they concluded that "deep structures" and "surface structures" were psychologically real, but the transformations relating them weren't.

© 2003 CSLI Publications

Early Evidence for the Psychological Reality of Deep Structures

- The proposed DS for (2) had three occurrences of *the detective*, while the proposed DS for (1) had only two:

  (1) *The governor asked the detective to prevent drinking.*

  (2) *The governor asked the detective to cease drinking.*

- In a recall experiment, *detective* was significantly more effective in prompting people to remember (2) than (1)

© 2003 CSLI Publications

Typical Problem Cases for the DTC

(1) *Pat swam faster than Chris swam.*
(2) *Pat swam faster than Chris did.*
(3) *Pat swam faster than Chris.*

- The DTC predicts that (1) should be less complex than (2) or (3), because (2) and (3) involve an extra deletion transformation.

- In fact, subjects responded more slowly to (1) than to either (2) or (3).

© 2003 CSLI Publications

What should a psychologically real theory of grammar be like?

- The "deep structure" distinctions that are not evident on the surface should be represented.
- The transformational operations relating deep and surface structures should not be part of the theory.
- Our information-rich trees include all of the essential information in the traditional deep structures, but without the transformations.

© 2003 CSLI Publications

Jerry Fodor claims the human mind is "modular"

"A module is…an informationally encapsulated computational system -- an inference-making mechanism whose access to background information is constrained by general features of cognitive architecture."

-- Fodor, 1985

A central issue in psycholinguistics over the past 20 years has been whether language is processed in a modular fashion.

© 2003 CSLI Publications

Tanenhaus's Eye-Tracking Experiments

- Participants wear a device on their heads that makes a videotape showing exactly what they're looking at.
- They listen to spoken instructions and carry out various tasks.
- They eye-tracking provides evidence of the cognitive activity of participants that can be correlated with the linguistic input.

© 2003 CSLI Publications

Non-linguistic visual information affects lexical access

- Participants' gaze settled on a referent before the word was completed, unless the initial syllable of the word was consistent with more than one object.
- For example, participants' gaze rested on the pencil after hearing *Pick up the pencil*
  more slowly when both a pencil and a penny were present.

© 2003 CSLI Publications

Non-linguistic visual information affects syntactic processing

- Eye movements showed that people hearing (1) often temporarily misinterpreted *on the towel* as the destination.

  (1) *Put the apple on the towel in the box.*

- When *on the towel* helped them choose between two apples, such misparses were significantly less frequent than when there was only one apple.

© 2003 CSLI Publications

General Conclusion of Eye-Tracking Studies

- People use whatever information is available as soon as it is useful in interpreting utterances.
- This argues against Fodorian modularity.
- It argues for a model of language in which information is represented in a uniform, order-independent fashion.

© 2003 CSLI Publications

Speakers know a great deal about individual words

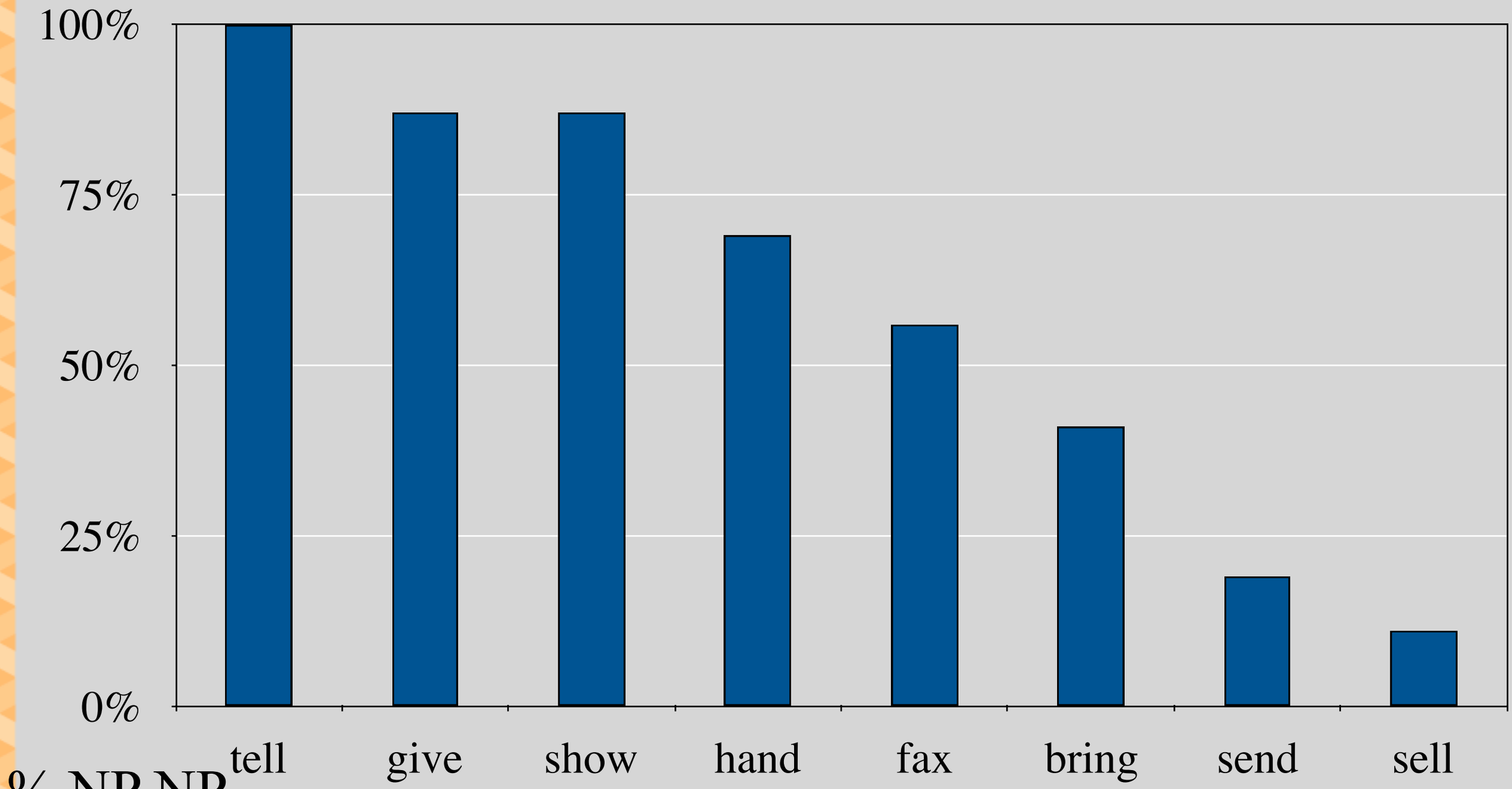
- Individual lexical items have many idiosyncrasies in where they can occur, and in where they tend to occur.

- For example, the verb *behoove* occurs only with the subject *it* (and only in certain verb forms), and the verb *beware* has only the base form.

- We also know that the transitive use of *walk* is much rarer than the intransitive.

© 2003 CSLI Publications

V-NP-NP vs. V-NP-PP Frequency in the *NYT*



© 2003 CSLI Publications

Lexical biases influence processing

- Wasow et al ran a production experiment to test whether ambiguity avoidance would influence speakers' choice between (1) and (2):

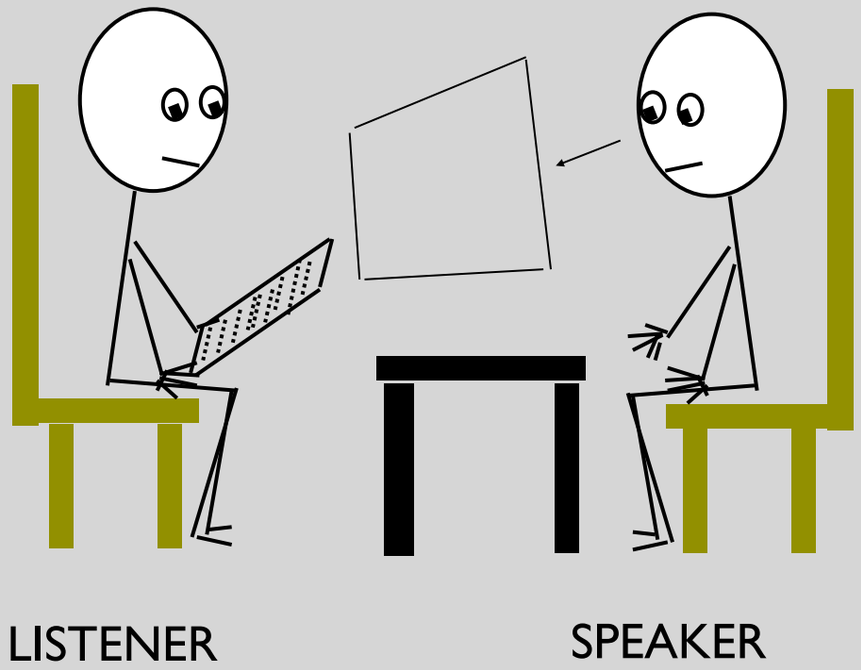
  (1) *They gave Grant's letters to Lincoln to a museum.*

  (2) *They gave a museum Grant's letters to Lincoln.*

- Lexical bias of the verbs turned out to be a significant predictor of which form speakers used (and ambiguity avoidance turned out not to be).

© 2003 CSLI Publications

Experimental Method

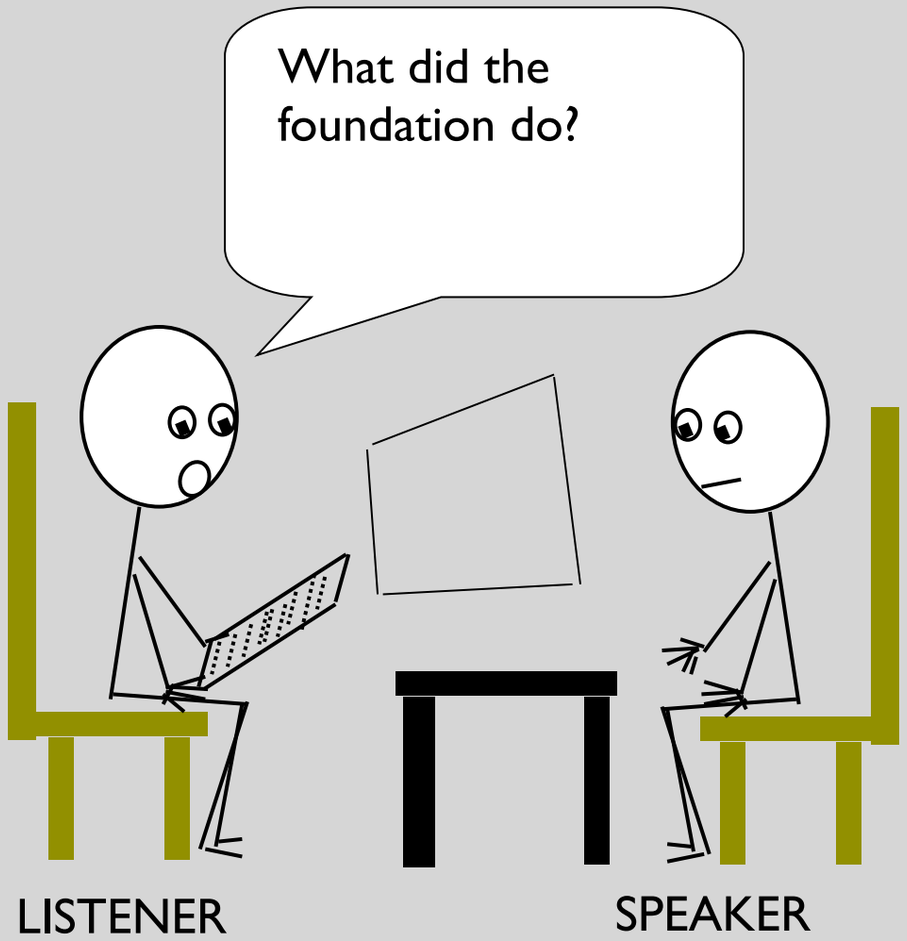


1. Speaker silently reads a sentence:

A museum in Philadelphia received Grant's letters to Lincoln from the foundation.

© 2003 CSLI Publications

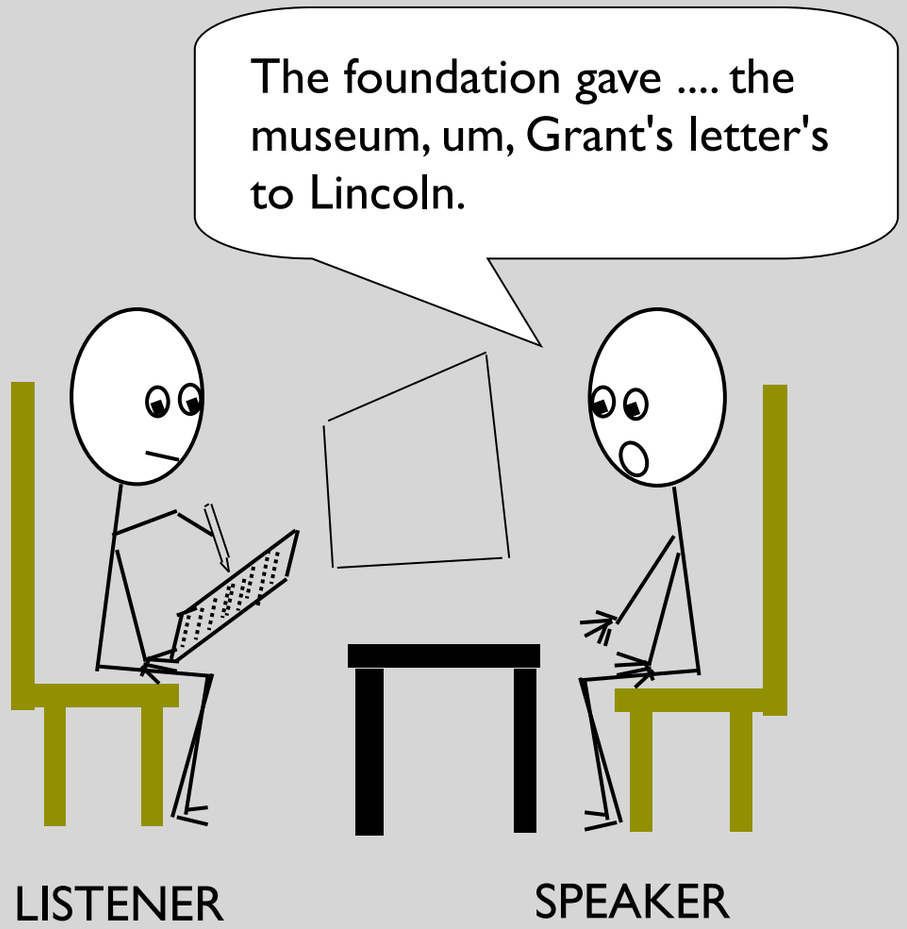
Experimental Method, continued



2. The sentence disappears from the screen.
   The listener reads the next question from a list.

© 2003 CSLI Publications

Experimental Method, continued

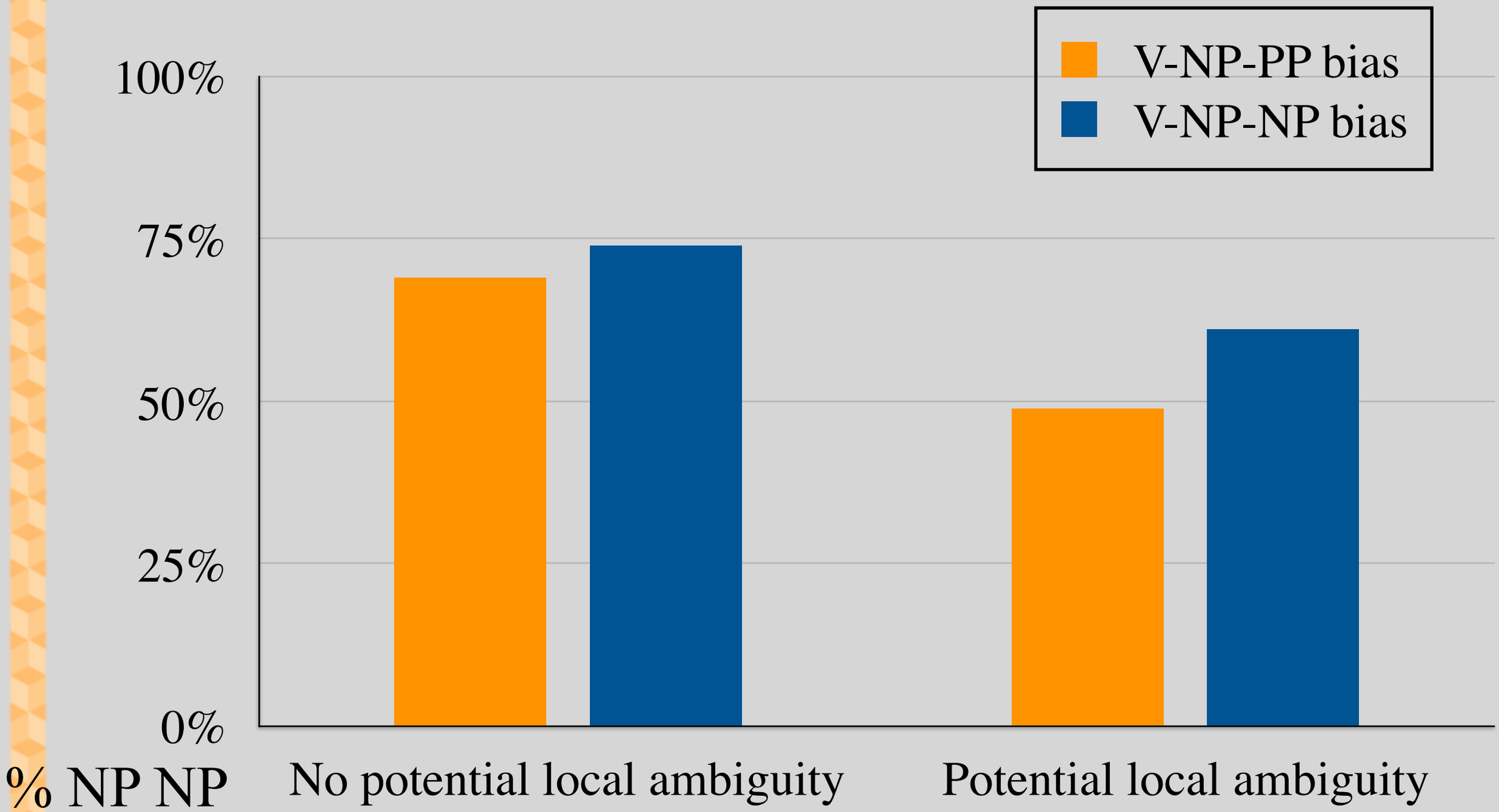


3. The speaker answers the listener's question.

   The listener chooses the correct response on a list (from two choices).

© 2003 CSLI Publications

Experimental Results on Local Ambiguity



© 2003 CSLI Publications

Reverse ambiguity effect

- Arnold, Wasow, Asudeh & Alrenga 2004 *Journal of Memory & Language*
- Re-ran the experiment with slightly better methodology and found a *stronger* reverse ambiguity effect.

© 2003 CSLI Publications

A psychologically real grammar should be lexicalist

- Early generative grammars downplayed the lexicon.
- Now, however, the importance of the lexicon is widely recognized.
- This aspect of grammar has been developed in greater detail in our theory than in any other.
- It would be easy to add frequency information to our lexicon, though there is debate over the wisdom of doing so.

© 2003 CSLI Publications

Conclusion

- Grammatical theory should inform and be informed by psycholinguistic experimentation.
- This has happened less than it should have.
- Existing psycholinguistic evidence favors a constraint-based, lexicalist approach (like ours).

© 2003 CSLI Publications

Universals?

- P&P (top-down): attempts to relate multiple typological properties to single parameters.
- Grammar Matrix (bottom-up(-ish)): attempts to describe many languages in a consistent framework and then takes stock of common constraints.

© 2003 CSLI Publications

Universals?

- Case constraint
- SHAC
- Binding theory
- Head-complement/-specifier/-modifier
- Head Feature Principle
- Valence Principle
- Semantic Compositionality Principle
- ...

© 2003 CSLI Publications

Overview

- Psycholinguistics and grammar design
  - What grammar has to say
  - What psychological evidence has to say
    - Acquisition
    - Production
    - Comprehension
- Universals

© 2003 CSLI Publications

Reading Questions

- Where do the grammar principles live, in our current version of the grammar? The SHAC is built into the type constraints. The Case Constraint is contained in the formal definitions in 9.2.7. But where are the HFP, the semantic principles, the ARP, the AAP, etc.? Just implied from previous chapters?

© 2003 CSLI Publications

Reading Questions

- I think I understand that HPSG, being constraint-based, is superior to the transformational grammar in that HPSG does not stipulate a fixed linear order, as the transformational grammar does with the deep structure.

© 2003 CSLI Publications

Reading Questions

- When p.303 talks about order-independence, they're talking about a grammar being able to support either utterance-->meaning (comprehension) or meaning-->utterance (production). Given our grammar's foundation of lexical trees, is the only difference between production and comprehension (as far as we're concerned) that production builds trees top-down and comprehension builds trees bottom-up?

© 2003 CSLI Publications

Reading Questions

- Does the fact that people process sentences word by word, using linguistic and nonlinguistic cues imply the lack of transformation? What's to say people are not simulating the possible structures, including the transformations as they hear sentences? I see how following the word order is more elegant, but don't see why it is so implausible that a transformation could be part of the simulations that listeners make as they process the sentence on-line. Is this because constraints are simpler computationally?

© 2003 CSLI Publications

Reading Questions

- This chapter compares HPSG with Transformational Grammar a lot. How does dependency grammar fits into this picture? Is it generative? What level of adequacy does it have?

© 2003 CSLI Publications

Reading Questions

- Since in our grammar, trees can only be constructed from expressions, not lexemes, we are required to make decisions about semantics/ syntax before we begin to construct a tree. So, how does our grammar handle the incremental processing of ambiguous sentences? Do we hypothesize that multiple trees are stubbed out using all possible words that match the constraints we know and then disqualified as the sentence progresses? And if there are multiple possible full parses do we use probabilistic information to choose one?

© 2003 CSLI Publications

Reading Questions

- Because language processing and production is certainly incremental, a strong lexicalist approach is a great way of judging acceptabilities in the face of language variation. But our SPR and COMPS only relate words to their immediate environments, could there be a need to find more lexical relationships in other languages? Such as a long-distance word interaction?

© 2003 CSLI Publications

Reading Questions

- In 9.3, does the "strongly" in "strongly lexicalist" refer foremost to most of the grammatical and semantic info being in the entries, or does it require that the entries correspond directly to words in sentences. Are both conditions necessary for strong lexicalism?

© 2003 CSLI Publications

Reading Questions

- For the three observations stated in 9.3, I would like to know the intuitions behind the third feature STRONGLY LEXICALIST. I am guessing it is because being lexicalist is more computational-friendly. Is it true that by contracting most information in the lexical entries, the grammar rules are easier to be processed by computers?

© 2003 CSLI Publications

Reading Questions

- The discussion in 9.5 reminded me of the importance of probability and contextual information in resolving ambiguity. For example, *Alice heard Bob on the phone* is an ambiguous statement that appears to have no guaranteed constituent structure unless we were to have understanding of the context. Would it be appropriate to leverage probabilistic or contextual information inside the grammar itself to solve this problem, or is this strictly a problem belonging to other domains?

© 2003 CSLI Publications

Reading Questions

- Hypothetically, what might be the best way to incorporate these non-linguistic elements that affect the interpretation of language into our grammar, such as tone, sarcasm, emotion, etc? Could we incorporate this into our SEM structure maybe? I feel like this, combined with probabilistic data, could help to resolve ambiguities in sentences in computational systems that are analyzing language. As a side note, I think if there's a good way to encode tone/ emotion into linguistic structures used in computation, we could create more "empathetic" acting machine conversationalists that are more sensitive. How do we also create a model that's robust and can accommodate differences in the grammars of different dialects of English while still getting the correct intended meaning, or even the imperfect grammar of young children?

© 2003 CSLI Publications

Reading Questions

- It was interesting to step back and think about how HPSG relates to theories of natural language since we have focused heavily on building a modeling system rather than exploring the nature of language. Has HPSG been used as a tool to contribute to or study any of these theories? I can imagine an interesting machine learning task to see how (or if) a computer derives the rules and principles of HPSG given sufficient input data or lexical entries.

© 2003 CSLI Publications

Reading Questions

- Are lexical rules a feature of all languages? What about isolating languages like Chinese?
- p.307 says phrase structure rules have linear ordering that is not hard to factor out to achieve universality. How? Do we need to devise some paramater-like concept to account for, for example, if specifiers come before or after the head?
- p.308 "Our central purpose in this book is to present a precise framework for the development of _descriptively adequate_ grammars for human languages." Why "descriptively adequate"? I recall that Carnie says it is only the second-best level of adequacy, inferior to "explanatorily adequate". Why do we not aspire for that?

© 2003 CSLI Publications